Stat100, vår 2018

**Løsning uke 16**

**Oppgave I**

**a)** Yij = i + ij, der ij ~ N(0, ) og uavhengige.

La Åge være okse nr. 1 o.s.v.

Yij være melkemengde fra datter nr. j av okse nr. i. i = 1,2, . .5. j =1, 2, . . . .8.

Yij = j + ij. er meningsløs fordi den ser på effekt av hvilket nummer kua får ikke av hvem som er faren.

i er gjennomsnittlig melkemengde (forventning) fra samtlige døtre av okse nr. i

ij er avviket mellom observert melkemengde og forventning for datter nr. j okse nr. i.

 er standardavvik i melkemengde for **alle** døtre av samme okse.

, gitt som gjennomsnittene i Tabell 1. kg

R2 = 0,24. 24 % av variasjonen i melkemengde kan forklares med far.

**b)** H0:1 = 2 = 3 = 4 = 5. H1: Minst to forventninger er forskjellige.

Test for en F verdi på 2,806. Dette er større enn tabellverdien på 5 % nivå med 4 og 35 frihetsgrader. Og dermed påstår vi at okse (kuas far) har effekt på melkeproduksjon.

**c)** Kontrasten: = fanger den etterspurte effekten.

H0: = 0 mot H1: : > 0

Estimate 230

Std. Error 108

t value 2.13

35 frihetsgrader, og er dermed har vi signifikant positiv effekt på 5 % nivå

(Tabellverdi t0,05,35 = 1,690).

**d)** Residualet blir: e1,1= 6363 – 6689 = -326 (kg. melk pr. år)

Negativt residual betyr at denne kua melker mindre enn gjennomsnittet av de 8 døtrene til Åge som er med I undersøkelsen.

e1,8 = 7562 – 6689 = 873. Denne kua melker mye mer enn snittet av de 8 døtrene til Åge som er med i undersøkelsen.

Minst residual er den dattera til Åge som melker minst.

**Oppgave 2 (Bruk R-commander)**

Finn gjennomsnitt og standardavvik i hver gruppe.

mean sd

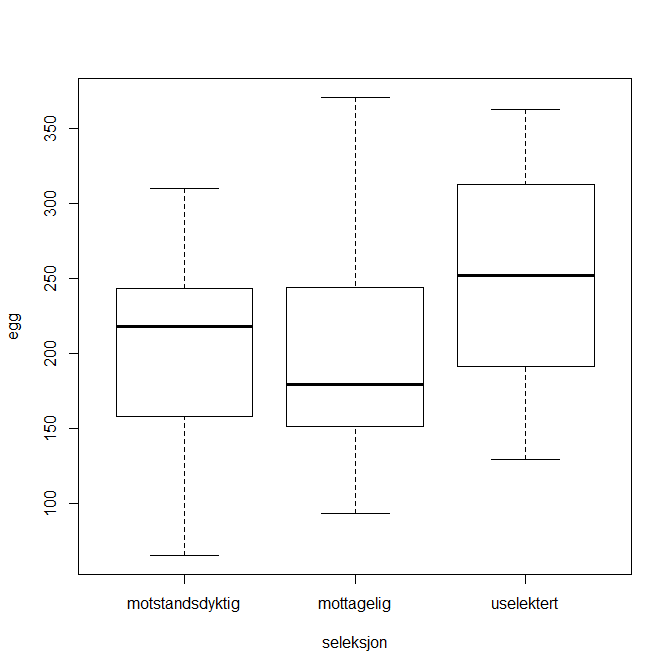
motstandsdyktig 205.48 65.11

mottagelig 194.84 65.69

uselektert 253.08 70.64

Ser det ut til at antagelsen om likt (populasjons)standardavvik er akseptabel?

Estimatene i hver gruppe er ganske like, det taler for at antagelsen ikke skulle være så gal.



Sett opp modellen for datanalysen du vil utføre og estimer alle parametere i modellen.

Yij = i + ij der ij ~N(0, ). Alle observasjoner er uavhengige.

i =1,2,3. j =1,2 . . . ., 25. Yij er antall egg for flue nummer j i gruppe nr i.

Kan du påvise signifikant forskjell mellom gruppene med hensyn på fruktbarhet?

Tester: H0: 1 = 2 = 3.

Mot alternativ hypotese: Minst to forventninger er forskjellig.

Df Sum Sq Mean Sq F value Pr(>F)

seleksjon 2 48091 24045 5.326 **0.00697**

Residuals 72 325079 4515

Sikker forskjell. Forkaster nullhypotesen.

**Oppgave 3** (bruk R-commander)).

Vi antar modellen der alle observasjoner er uavhengige.

Yij er poeng på brød j bakt med oppskrift i. i = 1,2,3,4, 5, 6 j =1,2, . . . ,8

Vi antar at poengene er uavhengige og at

**Yij = i + ij der ij ~N(0, ).**

H0: ingen forskjell mellom oppskrifter. D.v.s.

H0: 1 = 2 =3 = 4 =5 = 6 H1. minst to -er er forskjellige.

R Commander gir:

Df Sum Sq Mean Sq F value Pr(>F)

oppskrift 5 59.94 11.987 3.091 0.0183

Residuals 42 162.88 3.878

Denne testen har p-verdi 0,018 (hva betyr det?), og dermed kan H0 forkastes.

Vi er altså sikre på at oppskriftene er forskjellige, men hvor ligger denne forskjellen?

Mean SD

O1 6.625 2.263 8

O2 5.625 1.84 8

O3 7.875 1.552 8

O4 4.625 1.407 8

O5 4.750 2.375 8

O6 6.125 2.167 8

**Positiv effekt av grovt mjøl.**

 = (1 + 2 + 3)- (4 + 5 + 6)

H0:  = 0H1:  > 0

 1.54, SE() =0.57 T = 2.71 P verdi 0.009

Siden det er en ensidig test kan p-verdien halveres. Vi er sikre på at befolkningen foretrekker grovt mjøl.

***Effekt av nasjonalitet?***

 = (1 + 3 + 5)- (2 + 4 + 6) H0:  = 0 H1:  ≠ 0  0.958

SE() = = 0.568

Under H0 er T =  = 1,68 t-fordelt med 42 frihetsgrader. Sjekk med tabell, og du kan forkaste på 10 %. R-commander finner at p-verdien er 0,099.

Kan du forkaste nullhypotesen på 19 %? På 5 %?

For norske oppskrifter: Er grovbrød best?

 = 2 - (4 + 6) H0:  = 0 H1:  > 0

 0.25

SE() = = 0.85

Under H0 er T =  = 0.29 t-fordelt med 42 frihetsgrader. Du kan ikke forkaste H0.

Eventuelt direkt fra R-commander.

Estimate Std. Error t value Pr(>|t|) DF

0.25 0.852 0.293 0.7708 42

**Oppgave 4**

**a)** La Yij være fiskevekt for dam j med gjødseltype i, der i = 1, 2, 3, 4, og j = 1, 2, 3, 4, 5, 6.

Vi atar at vektene er uavhengige og at

Yij = i + ij, der + ij ~ N(, ). Alternativt Yij ~ N(i, ),

Vi velger denne modellen (enveis variansanalyse) fordi vi vil sammenligne mange grupper, men bare med hensyn på en faktor.

i er gjennomsnittlig fiskevekt dersom vi brukte gjødselslag nr. i på alle fiskedammer i hele populasjonen (f.eks. i hele landet)

 er standardavvikt i fiskevekt for alle dammer (i hele populasjonen) ved bruk av en type gjødsel.

**b)** Df Sum Sq Mean Sq F value Pr(>F)

gjødsel 3 5693 1897.5 26.95 3.15e-07 \*\*\*

Residuals 20

Vi tester

H0: ingen effect av gjødsel-slag

H0: 1 = 2 =3 = 4. H1: Minst to -er er forskjellige.

Denne testen har en p-verdi på 0,0000... og vi forkaster H0 (på alle nivå)

Alstså påstår vi at det har betydning hva slags gjødsel du bruker.

Dersom det ikke var noen effekt av gjødling med hensyn på fiskevekt, er sannsynlighet for å observere minste så store forskjeller mellom gjødseltyper som vi gjør i våre data, lik p-verdien.

**c, d og e)** La gruppe 1 vær kyllig-gjødsel, gruppe 2 være geitegjødsel, gruppe 3 være kunstgjødsel, gruppe 4 være ingen gjødsel.

Vi må sjekke følgende kontraster:

1 = (1 + 2 +3) - 4 2 = (1 + 2) - 3 3 = 1 - 2

Da er : 32,33 12,5 11

Standardfeielene (SE) til estimatene er 

SE() = 3,96 SE() = 4,2 SE() = 4,8

La H0 være = 0 for alle tre kontraster.

Under H0 vil T =  være t fordelt med 20 frihetsgrader.

T1 = 8,16 T2 = 2,98 T3 = 2,29.

Hvis vi f.eks. tester på 5 % vil alle kontraster være signifikantforskjellig fra eller større enn null. (sjekkes mot 2,08 tosidig og 1,72 ensidig).

Smartere er det å bruke R-commander. Etter at modellen er kjørt gå til Model > Test contrasts in ANOVA. Gi c-verdiene for hver kontrast. For den første får du denne utskriften

Estimate Std.Err t value Pr(>|t|) DF lower CI upper CI

gjødsel 32.333 3.95 8.17 8.33e-08 20 24.08 40.58